DOCUMENT RESUME

ED 104 903 95 TM 004 351

AUTHOR Cooley, William W.
TITLE Evaluation in a Research and Development Context.
INSTITUTION Pittsburgh Univ., Pa. Learning Research and Development Center.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE 73
NOTE 15p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, Louisiana, February 1973)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS Curriculum Development; *Educational Development; Educational Improvement; *Educational Research; *Evaluation; Evaluation Methods; *Interaction; *Models; Psychological Studies; Research and Development Centers; Validity; Values
IDENTIFIERS Valuation Theory

ABSTRACT

Educational research and development (R&D) has often been characterized as a neat, linear sequence of discrete steps, moving from research through development to evaluation and dissemination. Although the inadequacies of such linear models of educational research and development have been pointed out previously, these models have been so much a part of R&D thinking that it is useful to point them out again. In addition, the attributes of an interactive model are shown to surpass the linear model and to enhance the description of the R&D process. This model is discussed in the context of the Learning Research and Development Center (LRDC) where interactions take place among basic and applied research, development, evaluation, and dissemination. It is also discussed in conjunction with the theoretical requirement of John Dewey's theory of valuation which states that the value which people will attribute to a new product or procedure will be a function of how well they think it will satisfy some recognized needs. The interplay of this interaction model and valuation theory is presented. (Author/BJG)

ED104903

EVALUATION IN A RESEARCH AND DEVELOPMENT CONTEXT

William W. Cooley

Learning Research and Development Center

University of Pittsburgh

1973

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TM 004 351

Published by the Learning Research and Development Center, supported in part by funds from the National Institute of Education (NIE), United States Department of Health, Education, and Welfare. The opinions expressed do not necessarily reflect the position or policy of NIE and no official endorsement should be inferred.

EVALUATION IN A RESEARCH AND DEVELOPMENT CONTEXT[1]

William W. Cooley

Center for Advanced Study in the Behavioral Sciences[2]

Educational research and development has often been characterized as a neat, linear sequence of discrete steps, moving from research through development to evaluation and dissemination (e.g., Clark & Guba, 1965). Although the inadequacies of such linear models of educational research and development have been pointed out previously (see, for example, Baldridge & Johnson, 1972), these models have been so much a part of R&D thinking that it is still useful to illustrate their shortcomings. Consider this scenario: The basic researcher, A, makes a psychological discovery regarding cognitive processes involved in comprehension by novice readers. B, the applied scientist, sees the possible implications of this discovery, and translates A's findings into validated operating principles with implications for instructional psychology. C, the developer, takes these new principles and builds a new reading program. D, the evaluator, field tests this product, finds

---

[1] This paper was part of a symposium on Interactive Psychological Research and Curriculum Development presented at the meeting of the American Educational Research Association, New Orleans, February 1973. Portions of the paper are based on a book being prepared with Paul R. Lohnes entitled Evaluative Inquiry in Education.

[2] On leave from the Learning Research and Development Center, University of Pittsburgh, 1972-73.

that it achieves C's objectives, and gives it a seal of approval. E, the publisher, develops the training and sales programs necessary for dissemination. Finally, teachers and kids, the consumers, live happily ever after! Although this scenario illustrates the linear model for R&D, it certainly does not resemble what takes place in educational research and development.

An interactive model provides a better description of the R&D process. In an R&D environment, such as the Learning Research and Development Center (LRDC), interactions take place among basic and applied research, development, evaluation, and dissemination. Notice that the interactions are described as being among activities, not among distinct types of people. This is because individuals do not restrict their attention to only one narrow aspect of R&D, but become involved in different aspects of the R&D process. Much more could be said about the general nature of the interactions between research and development in a setting such as LRDC, but this paper deals with a more specific topic: evaluation and its interactions with psychological research and curriculum development. Examples of such interactions are drawn from the work being done at LRDC on a New Primary Grades Reading System (NRS)(Beck & Mitroff, 1972).

Because "evaluation" has come to mean quite different things to different people, it is first necessary to clarify the mission of the set of activities called evaluation at LRDC. Evaluation involves the collection and organization of data into information relevant to judgments of value, the objects of these judgments being the products and procedures which LRDC is developing for use in the schools. Evaluation activities are called evaluative research in recognition of the fact that the generation of such information or evidence, if it is to be believed by

those for whom it is being prepared, must adhere to the same sound research principles which guide any disciplined inquiry. What distinguishes evaluative research from basic research is whether the propositions with which it deals are related to policy or to theory. Evaluative research investigates the validity of propositions which involve means and ends and their relationship to educational policy or practice, rather than propositions about cause and effect and their relationship to psychological theory.

This is not to say that effective evaluative research is theory-free. Theory certainly is relevant, at least in the sense that organizing ideas are essential to guide the work. But in evaluation, the questions to be investigated do not derive from theory. They derive from the world of action, and the results produced are fed back into that world (see, for example, Coleman, 1972 for further discussion of this point). One theoretical requirement is a theory of valuation. A very serviceable theory was provided by John Dewey. A brief summary of this theory and its implications for the evaluation of educational systems is presented below.[3]

Quite simply, the theory says that the value which people will attribute to a new product or procedure will be a function of how well they think it will satisfy some recognized needs. LRDC's mission is to improve education by providing schools with more effective products and procedures than are currently available. To have such impact, schools must implement what LRDC develops. They will do this only if

---

[3] A more complete summary can be found in Dewey (1939) and in Evaluative Inquiry in Education, a forthcoming book by the author and Paul R. Lohnes.

they see the need for it and are convinced that the product is demonstrably effective in fulfilling the need. The function of evaluative research is to provide information which demonstrates the need and shows how a particular innovation is a means to satisfy that need. Nothing startlingly new here, but if these value concerns are not kept in mind, the evaluator will tend to stop short of the total task, and, for example, show only that NRS does in fact teach the skills it was designed to teach. This latter demonstration is a necessary part of the evaluation task, but it is not sufficient for determinations of value.

The specified need defines the end-in-view, and a means-ends continuum is proposed which establishes the relation between the innovative means and that end-in-view. The relative importance of the need and the credibility of the means-ends propositions, as seen by the valuer, will determine how highly the innovation will be valued by that valuer. With respect to a new reading system, one might try to make the case that it will increase grade equivalent scores on reading ability by six months for children from lower socioeconomic level families. If such an increase brings the mean for these children up to grade level, then the reading system will be valued by those who viewed grade-level achievement as a desired end.

There may be more people, however, who are worried about the 21 million adults, products of our schools, who are unable to read well enough to fill out a job application, vote, apply for a loan, obtain information on such basics as medical aid, etc. If one could show the relationship between the use of NRS and the reduction of functional illiteracy in this country, then perhaps NRS would be valued even more highly.

One might try to go even further and show how NRS could have been an effective weapon in the war on poverty, but such an effort would have been in vain according to Jencks, Smith, Acland, Bane, Cohen, Gintis, Heyns, and Michelson (1972), because their evaluative research suggests that the means-ends propositions which link the improvement of schooling with the reduction of poverty are not valid. But the point is that the greater the established need and the more convincing the relationship between the new means and the reduction of that need, the more highly the new means will be valued.

Thus, one form of interaction among evaluation, research, and development is in the determination of need. Need is not demonstrated by a small-scale market survey to find out what consumers will buy next year, but rather by an analysis of critical educational and societal problems with the expectation that if this analysis guides development work, what gets "built" will more likely get used. Such analyses would include a consideration of the major educational needs expressed by the critics of today's schools, parent-community committees on educational goals, and national educational policy groups, as well as consideration of the implications of empirical work of national, state, and local assessments, Harris polls (which publicized the illiteracy problem mentioned above), etc.

The other aspect of the valuing process is believing that the new means will move you toward the desired end-in-view. People have to believe something will work before they will buy it. This belief is a function of their seeing <u>how</u> it works as well as seeing <u>that</u> it works. Although buyers don't insist on understanding how a color TV set works before they buy it, they are more willing to believe consumer satisfaction data if it is accompanied by some plausible functional difference

between one particular set and other sets on the market. Therefore, another aspect of the evaluation of NRS is making explicit its main features and the learning principles upon which it is based, and how they differ from alternatives. This, of course, is not necessarily done by someone with the name "evaluator" on the door, but is another example of the necessary interaction among evaluation, research, and development activities.

It is also important to recognize that valuation theory makes it clear that, given a means-ends continuum, value can be attached to both means and ends. Also, everyone "knows" that the ends do not justify the means. Thus, evaluation cannot restrict its attention to outcomes. There is no sharp distinction between product and process, and both are subject to valuation. In fact, many parents are more concerned about the kinds of experiences their children have in school than about the specifics of what their children are learning. One implication of this point for developers is the importance of making it clear to others what it is like for a child to be in their program.

The most demanding evaluation research task is generating information which clearly shows the learning effects of a new program, not just in the laboratory, or in a carefully controlled developmental school, but also in the field, in schools representative of those to which the evaluation results are to be generalized. Field testing is complicated because of the variety of factors which affect the outcomes of an educational system and because any program as intricate as NRS will surely get implemented in the field in a variety of ways.

The variation in implementation which inevitably takes place from classroom to classroom provides an important focus for interactions among evaluation, research, and development. To discuss this

point, it is useful to refer to Figure 1, which suggests that what a child learns during a given time interval (e.g., during first grade) is a function of instruction, family, and peer group differences, and, most importantly, of the abilities with which the child began this time interval (T_1 abilities). It has most certainly been established that each of the four sets of variables uniquely explains portions of the variance in T_2 abilities. Thus, all four must be considered in evaluation studies which attempt to identify instructional differences which affect how well children learn.

This evaluation model does not imply pitting NRS classrooms against brand X classrooms and contrasting the two means. Very little is learned from such exercises. One certainly cannot assume that a given method will be uniformly implemented in the classroom. Also, any two reading programs will initially be alike in some respects and different in others. Reading programs in a sample of first-grade classrooms (with or without NRS) might differ in the following ways:

1. Degree of learner control (e.g., how much of the program is prescribed, how much is exploratory?)
2. Nature, frequency, and purpose of testing (e.g., how often are diagnostic tests individually administered?)
3. Individualization in goals (i.e., is goal setting done on an individual basis?)
4. Individualization in rate (i.e., are children permitted to progress at different rates?)

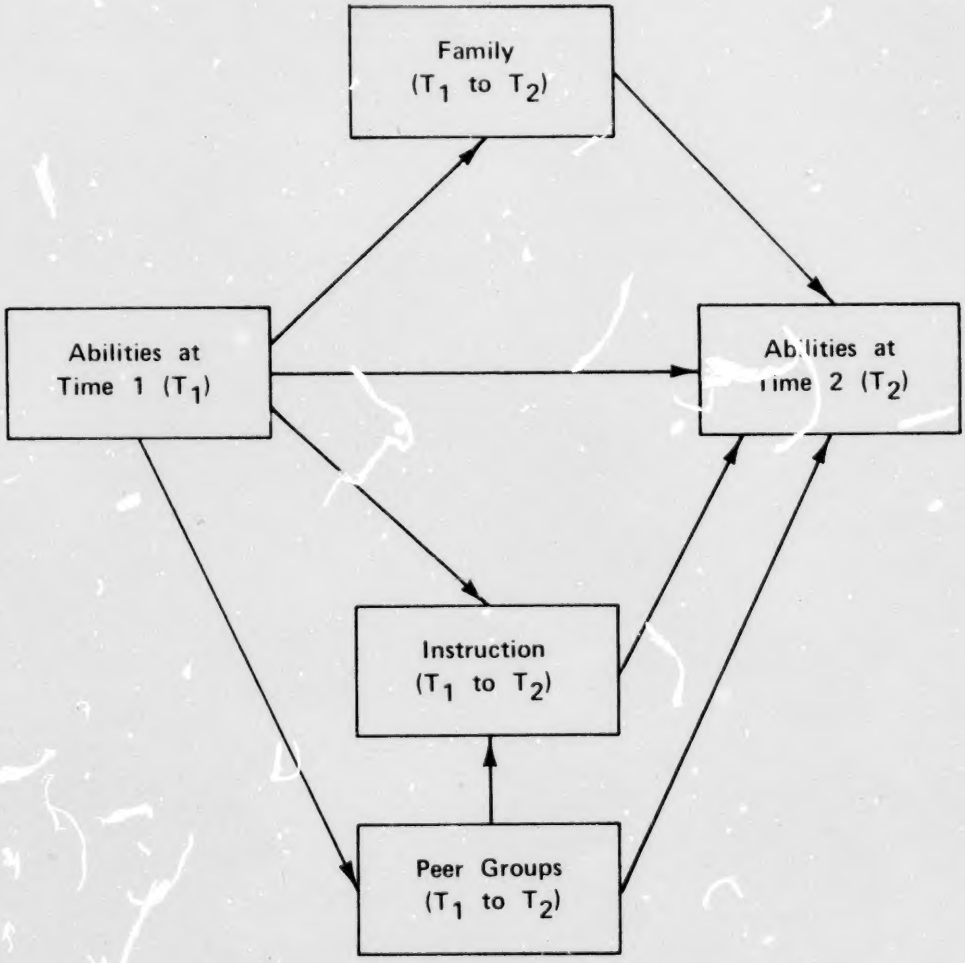

Figure 1. Functional Relationships Among Sets of Variables

5. Teacher behaviors (e.g., frequency and distribution of contacts with individual children; percentage of contacts which are negative)

6. Control of time (e.g., flexible or rigid; variable from day to day or from student to student depending upon need)

7. Variety and quality of available instructional materials (e.g., are games and manipulables used in the classroom?)

8. Mastery learning (i.e., is level of achievement required before moving on to the next lesson?)

9. Code-breaking approach (e.g., degree to which grapheme/phoneme relationships are taught directly and practiced in isolation)

Specifications for the measurement of instructional differences that are likely to make a difference in student achievement are critically dependent upon successful interactions among specialists concerned with evaluation, research, and development, and the results have implications for all three. Examples from a recent dissertation by Leinhardt (1972) illustrate this point.

The setting for her research was second-grade classrooms from LRDC's two developmental schools and from four school systems participating in the national Project Follow Through. Leinhardt developed measures of several kinds of implementation variables, and made observations in 30 classrooms, all of which were using some form of LRDC's

instructional program. She found that her measures of implementation were able to explain a considerable portion (46 percent) of the variance in T_2 abilities that was not explained by T_1 abilities. An important point here is how different this finding is from the kinds of results which led Jencks et al. (1972) to the conclusion that school differences don't make a difference. They may be right, in part because about two-thirds of the variance in output abilities is eliminated when school is used as the unit of analysis, and in part because school variables (e.g., cost/pupil) only indirectly affect instruction. School differences may not make much of a difference, but classroom differences may be very important!

Looking beyond the variance explained by the instructional variables, to their structure as predictors, it is possible to see the implications of this approach to evaluation for instruction theorists, developers, evaluators, and consumers. For example, the Leinhardt data revealed a surprising relationship between frequency of testing and amount of learning which had taken place during the year. It suggested that teachers who overrode the system and tested less frequently than LRDC recommended, showed higher ability gains for their classrooms than did the teachers who tested more frequently. If the evaluator can convince the developer that this is a valid finding, then modifications can be made in the instructional program in its next approximation. If the instructional theorist decides that the finding is a valid one, it may have implications for his (her) instructional model, at least the part about the necessity of constant monitoring of student progress once placement of the student has been achieved.

It is the consumer who will benefit the most if evaluations describe instructional effects in terms of multidimensional contributions

instead of gross contrasts of pre-package plans (with no substitutions please!). Of course, it will take some getting used to in consumer-land. Buyers tend not to be provided information which tells them that they could add a $5 gadget to their TV set and have a picture of the same quality as a new $500 set would provide. But that is the kind of result our evaluation model tends to produce.

The typical contrast of gross treatments (e.g., NRS vs. Distar) usually results in a no-significant-difference finding, from which very little is learned. One reason for the no-difference finding may be that although both developers made important contributions, they affected different aspects of the reading problem, with the same net effect for the two reading systems. Another possibility is that one of the systems is really superior (i.e., based upon sounder principles) but was unevenly implemented, thus losing its potential impact. But the most important reason for moving from an analysis of variance mode, which contrasts gross treatments, to a multidimensional model of instructional differences, has to do with the mission of LRDC.

The mission of this R&D center is to improve education in elementary schools through an improved understanding of the instructional process. An evaluation model which can contribute to those understandings is far more appropriate than one designed to promote products which mysteriously produce magical effects.

REFERENCES

Baldridge, J. V., & Johnson, R. The impact of educational r&d centers and laboratories. Stanford: Stanford University, 1972.

Beck, I., & Mitroff, D. D. The rationale and design of a primary grades reading system for an individualized classroom. Pittsburgh: Learning Research and Development Center, 1972. (Publication 1972/4)

Clark, D. L., & Guba, E. An examination of potential change roles in education. Washington, D. C.: National Education Association Center for Study on Instruction, 1965.

Coleman, J. S. Methodological principles governing policy research in the social sciences. Paper presented at the meeting of the American Association for the Advancement of Science, Washington, D. C., December 1972.

Dewey, J. A theory of valuation. In International encyclopedia of unified science. Vol. II, No. 4. Chicago: University of Chicago Press, 1939.

Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H., Heyns, B., & Michelson, S. Inequality: A reassessment of the effect of family and schooling in America. New York: Basic Books, 1972.

Leinhardt, G. The boojum of evaluation: Implementation, an instrument. Unpublished doctoral dissertation, University of Pittsburgh, 1972.